Longitudinal nonresponse in the Current Population Survey (CPS)

Harris-Kojetin, Brian A.; Tucker, Clyde

Veröffentlichungsversion / Published Version
Konferenzbeitrag / conference paper

Longitudinal Nonresponse in the Current Population Survey (CPS)[1]

BRIAN A. HARRIS-KOJETIN AND CLYDE TUCKER

Abstract: *A longitudinal database was created consisting of seven complete "cohorts" who were in the CPS sample all eight months over nearly a 2 year period. Comparisons were made among complete respondents and partial respondents. Households that were complete respondents were relatively more likely than households that were partial respondents to be located in rural areas and in any region of the country except the West, and to have more persons living there, and to be occupied by owners. Members of complete respondent households were relatively more likely than members of partial respondent households to be a married couple, children, white, non-Hispanic, over 65 years of age, and were relatively more likely to answer a question on total family income. Persons in partial respondent households had a higher level of employment in the first four months-in-sample (MIS) and a higher unemployment rate in MIS 1 than persons in complete respondent households.*

Keywords: *panel nonresponse.*

1 Introduction

It is a fact of modern life that organizations that survey households, individuals, or business establishments will not obtain complete cooperation from everyone they solicit. Survey nonresponse occurs at several different levels, with individuals or households not responding at all (unit nonresponse), not responding during one wave or panel of a longitudinal survey (wave/panel nonresponse), or simply omitting certain survey items (item nonresponse). Ideally, one would hope that nonrespondents are a random cross-section of the sample, reflecting the same demographic, geographic, and economic groups as respondents. If this is the case, then one need not be concerned about obtaining biased results from a survey in which there was some degree of nonresponse. However, it is typically the case that nonrespondents differ from respondents on these characteristics (for reviews see Goyder 1987, Groves 1989). Therefore, even surveys with high response rates may have some degree of bias in their results to the extent that nonrespondents differ from

[1] The views expressed in the paper are those of the authors and do not necessarily represent those of the U.S. Bureau of Labor Statistics.

the respondents.

Of course, it is the nature of nonresponse that we almost never know exactly what we wanted to know about survey nonrespondents. Longitudinal surveys or those with a panel component can provide insights into some of the characteristics of some nonrespondents who provide information on a previous or subsequent panel or wave. Longitudinal surveys spanning several years are particularly concerned with bias entering due to attrition (e.g., Zabel 1994), whereas shorter panel surveys may be more likely to have people missing a wave and then returning. With panel nonresponse, if one can assume that the characteristic of interest was stable during the entire time then little is lost by missing one observation in time, because the data can be drawn from another point in time. However, most panel surveys are primarily concerned with characteristics that are (at least potentially) changing over time and, therefore, observations from other points in time provide a clue as to what the actual response would have been, but may not be sufficient to infer accurately the desired characteristics.

1.1 The present study

The focus of the present study is on panel nonresponse in the Current Population Survey (CPS). We created a longitudinal database file for the CPS consisting of a number of "cohorts" who were in the CPS sample for the full 8 months over a 16 month period. In this paper, we examine patterns of longitudinal nonresponse and the characteristics of nonrespondents using information obtained in months in which there was an interview. Previous research on panel nonresponse to the CPS has been severely hampered by the nature of the control card data and the difficulty of tracking individuals and households accurately across time. For example, one prior study was able to match only 50% of the cases across all 8 panels of the CPS (Dippo et al. 1992). Also, it was not possible to distinguish movers from other unmatched households. Beginning in January, 1994 the redesigned questionnaire of the CPS was implemented with new computer-assisted data collection methods which included among several improvements in data quality, the creation of a longitudinal identification number that would allow better longitudinal record-level matching of CPS data.

2 Design

The CPS is the monthly household labor force survey for the United States conducted by the U.S. Census Bureau for the U.S. Bureau of Labor Statistics. Approximately 60,000 eligible households are sampled each month in a two-stage clustered design. Households selected for sample are interviewed for 4 consecutive months, are not interviewed 8

months, and then are interviewed again for 4 consecutive months. Furthermore, in any given month, one eighth of the sample is composed of households participating for the first time (month-in-sample 1 (MIS 1)), one eighth the second time (MIS 2), etc.

Data for the present investigation were drawn from the first seven cohorts that completed all of the eight months-in-sample (MIS 1-8) since their initial selection during or after January 1994. The first cohort began in January 1994 and finished in April, 1995, and the last cohort began in July, 1994 and completed in October, 1995. This data set includes a total of 75,854 households.

2.1 Matching procedure

Household level matching. The longitudinal matching involved several levels of analysis because nonresponse is at the household level whereas survey response is at the person level. First, a longitudinal household level data file was created. A total of 60% of the households matched across all eight MIS. The majority of the non-matching cases were out of scope at least one month (e.g., demolished, nonresidential use, vacant, etc.) or were movers. Of the 45,395 households that matched across all eight MIS, 35,018 had some combination of interviews or Type A nonresponses (i.e., refusals, noncontacts, other noninterviews) for all eight MIS and are the focus of analyses in this paper. Of those, 28,724 households had interviews only for all eight MIS, 699 households had nonresponses only for all eight MIS, and 5595 households had at least one interview and at least one nonresponse during their eight MIS.

Person level matching. For the present paper, only those households with interviews all eight months or at least one interview and at least one nonresponse were selected for examination at the person level. A total of 99,639 persons were in those households, with 84,665 persons in households that were interviews all eight MIS, and 14,974 persons in households that were in all eight MIS and had at least one interview and at least one nonresponse.

2.2 Measures

The measures included in the present investigation covered several different levels of analysis. Geographic area information available included region of the country, urbanicity, and whether the area had a high percentage of people living below the poverty line. These data were available for all households regardless of their participation in the CPS. Additional data at the household and person levels was available only for those households with at least one interview. Household-level information included whether the

housing unit was occupied by the owner or a renter, the total number of persons in the household, the type of family (e.g., couple with children, single parent, single adult), and total family income. Person-level information included demographic characteristics such as age, sex, race, ethnicity, education level, marital status, and household members' relationships with each other. For the present study, the only substantive survey information examined was labor force classification (employed, unemployed, and not in the labor force), which is the primary focus of the CPS.

2.3 Analysis plan

Analyses comparing respondent and nonrespondent cases were carried out at both the household and person level. Analyses were conducted to compare characteristics of households that were interviewed all eight months (complete respondents) with those interviewed at least once with one or more Type A nonresponses (partial respondents) on geographic, household, and person characteristics that were available only for cases which had been interviewed at least once. The geographic analyses also included households that were nonrespondents all eight MIS (complete nonrespondents). All available geographic, household, and person-level information was entered into a logistic regression to compare, as completely as possible, persons who were complete respondents to partial respondents. Initial comparisons of the labor force status of complete and partial respondents were also made. All analyses were conducted using unweighted data.[2]

3 Results

3.1 Longitudinal response patterns

Approximately 82% of the households were interviewed all eight MIS. The most common pattern involving Type A nonresponse was for the first month to be a nonresponse and all subsequent months to be interviews (2.8% of all the households) and the second most common pattern was nonresponse during all eight MIS (2.0% of all households). The other most common patterns most frequently involved one nonresponse and seven interviews. Patterns that showed attrition from the sample made up a total of 3.1% of the households.

[2] All significance tests reported in this paper use unweighted data and do not take into account the complex sampling design of the CPS.

3.2 Household level analyses

3.2.1 Geographic area characteristics

To examine whether there were significant differences in geographic area characteristics among households that were complete respondents, partial respondents, and complete nonrespondents, distributions of characteristics for these three groups were compared using chi-square tests. Significant differences were revealed between these three groups for region of the country, urbanicity, and poverty of the area (all p's $< .01$). The largest differences were between the complete respondents and the complete nonrespondents. There were a relatively greater proportion of households that were complete nonrespondents in the Northeast, West, and South, in urban areas (especially central cities), and in low poverty areas, while a relatively greater proportion of households were complete respondents in the Midwest, in rural areas, and in high poverty areas.

3.2.2 Household characteristics

Comparisons were also made at the household level for characteristics of the households that were obtained in the interview, but this information was not obtained for households that were complete nonrespondents. Partial respondent households were relatively more likely than complete respondent households to be occupied by renters, have fewer persons living there (1 or 2), and to be headed by a single parent or single person (all $p's < .01$). In addition, members of partial respondent households were relatively less likely to answer a question on total family income, while members of complete respondent households were relatively more likely to provide family income data or have a higher family income than partial respondent households.

3.3 Person level analyses

3.3.1 Demographic characteristics

There were also significant differences between complete and partial respondents on a variety of demographic characteristics. Households that were partial respondents were relatively more likely to contain persons who were nonwhite, Hispanic, who had graduated college, and who were 25-34 years of age; while households that were complete respondents were relatively more likely to contain persons who were white, non-Hispanic, under 19 years old, and who have never been married.

3.3.2 Multivariate analyses of demographic, household, and geographic variables

To take into account all of the geographic, household, and demographic characteristics in comparing complete respondents to partial respondents, logistic regression analyses were conducted entering all of the variables into the equation in order to maximally distinguish between these two groups. The overall difference between complete respondent households and partial respondent households was significant, $p < .01$, and the odds ratios are given in Table 1. The results are essentially the same as the univariate results reported above with most of the characteristics independently predicting interview status, except the effects of poverty and marital status are no longer significant. Although the overall equation was significant, even all together these variables were able to account for only a small percentage of the variance between these two groups, $R_L^2 = .040$.[3]

3.3.3 Labor force classification

Unlike most of the demographic characteristics noted above, a person's labor force classification may change from month-to-month during the time that they are in the CPS sample; therefore, one cannot determine with certainty a person's labor force classification during the months the person was a nonrespondent. Accordingly, comparisons of labor force classification were made for each MIS separately between the complete respondents and partial respondents. Only persons who were interviewed in a particular MIS and had a labor force classification are included. Therefore, the persons who were partial respondents and who had a labor force status in MIS 1 did not respond at some later month, while the persons who were partial respondents who had a labor force status in MIS 8 did not respond during some previous month. The distribution of labor force classifications for persons in complete and partial respondent households showed significant differences (p's $< .05$) in MIS 1-4, demonstrated marginally significant differences for MIS 5 & 6 (p's $< .10$), but were not significantly different in MIS 7 & 8. Persons in partial respondent households were relatively more likely to be employed during MIS 1-4 than persons in complete respondent households. The unemployment rate for persons in partial respondent households was also significantly higher ($p < .05$) than the rate for persons in complete respondent households during MIS 1, but the differences in the other MIS were not significant.

---

3 $R_L^2 = \dfrac{-2\log L_0 - (-2\log L_1)}{-2\log L_0}$

Table 1: Results of logistic regression predicting partial respondents (compared to complete respondents) with geographic area, household, and demographic characteristics

Variable	Odds Ratio
Region (West omitted)	
Northeast Region	.93*
Southern Region	.90**
Midwest Region	.66**
Urbanicity (rural omitted)	
Central City	1.25**
Balance of MSA	1.08**
Poverty area	1.04
Own Residence	.90**
Household Size (5+ omitted)	
1 Person Household	1.96**
2 Person Household	1.79**
3 Person Household	1.42**
4 Person Household	1.21**
Family type (singles omitted)	
married couple family	.90*
single parent	1.02
Refused/DK income	4.06**
Sex (1=male, 2=female)	.97
White/nonwhite	.75**
Hispanic/nonhispanic	.90*
Age (65 + omitted)	
16-19 years	2.41**
20-24 years	2.06**
25-34 years	1.93**
35-44 years	1.80**
45-54 years	1.62**
55-64 years	1.43**
Education (College degree + omitted)	
Less than High School degree	.85**
H.S. diploma	.95
some college	.91**
Marital status (never married omitted)	
Married	1.00
Widowed/Divorced/Separated	.97
Relationship (non-relative omitted)	
Reference person	1.07
spouse of reference person	1.06
child	.72**
other relative	.99

** $p < .01$, * $p < .05$

4 Discussion

4.1 Summary of findings

The present study found a variety of differences in geographic, household, and person characteristics between households that were complete respondents and those that were only partial respondents, i.e., households that had interviews and at least one nonresponse. Specifically, households that were complete respondents were relatively more likely than households that were partial respondents to be located in rural areas and in any region of the country except the West, and to have more persons living there, and to be occupied by owners. Members of complete respondent households were relatively more likely than members of partial respondent households to be a married couple, children, white, non-Hispanic, over 65 years of age, and were relatively more likely to answer a question on total family income. Although there were many significant differences between complete and partial respondent households on these characteristics, they actually accounted for a rather small proportion of the variance between the groups.

Differences in labor force classification were also demonstrated between persons in complete and partial respondent households. Persons in partial respondent households had a higher level of employment in the first four MIS and a higher unemployment rate in MIS 1 than persons in complete respondent households. Furthermore, the differences between these two groups were the most extreme at MIS 1 and decayed over the months in sample until there were no differences in MIS 7 & 8.

4.2 Implications

The evidence that households that are nonrespondents in some MIS are more likely to be from distinguishable geographic, demographic, economic groups leads to concern. The present results have possible implications for post-survey nonresponse adjustment procedures as well as field procedures. Current CPS nonresponse adjustment procedures do not utilize prior information obtained from the household. Of the variables examined in the present investigation, response adjustment utilizes only MSA status. Clearly, now that accurate longitudinal data on households and persons is easily accessible, there are many more possibilities for enhanced and refined nonresponse adjustment and imputation. The current results suggest efforts in this area may be quite worthwhile.

Very little information is currently obtained in the field about nonrespondents. Changes in field procedures that put more emphasis on obtaining more characteristics of the household and persons living there is essential in furthering research and understanding of nonresponse. The households that were complete nonrespondents all 8 MIS appeared to be distinctly different from households that were complete respondents as well as

households that were partial respondents. However, there were very few available characteristics on which comparisons of these households could be made. To the extent that complete nonrespondents are distinctly different from partial nonrespondents, our ability to make more accurate nonresponse adjustments will be limited by the information that is obtained in the field.

Further analyses need to be conducted on the labor force status of persons in partial respondent households. There is some indication of differences in labor force status of persons who become nonrespondents in later MIS (see also Tucker and Harris-Kojetin, pp. 45-54, this volume). Whether this leads to a bias in labor force estimates requires further investigation.

4.3 Limitations

Because the data used in this study were from a limited time period, the impacts of seasonal effects and real economic change could not be assessed. These factors also may have influenced our findings. All analyses presented in this paper have been based on unweighted data. Therefore, conclusions pertaining to rates and percentages do not necessarily represent the entire U.S. civilian noninstitutional population.

4.4 Future directions

The present findings should be considered preliminary as this is the first in what we expect will be an ongoing line of research on the longitudinal aspects of the CPS. Further analyses will also examine the demographic characteristics associated with different patterns of nonresponse and reasons for nonresponse (refusal, noncontact, other). The use of unweighted data is a good starting point in this type of research, however, to make any geographic conclusions or any others that would reflect the nation, base weights, which reflect the probability of selection, need to be used. The baseweighted data has not been adjusted for nonresponse and thus would not exaggerate nonresponse results (whereas final weighted data would). Evaluations of the current weighting procedures are planned, and additional analyses of the effect of methodological factors such as mode of interview on response are also planned.

References

Dippo, C. S., Silberstein, A. and Rope, D. (1992). An Examination of Longitudinal Nonresponse in the CE and CPS. Paper presented at the Third International Workshop on Household Survey Nonresponse, Voorburg, The Netherlands

Goyder (1987). The Silent Minority: Nonrespondents in Sample Surveys. Boulder, CO: Westview Press

Groves, R. M. (1989). Survey Errors and Survey Costs. New York: Wiley

Tucker, C. and Harris-Kojetin, B. (1997). The impact of nonresponse on the unemployment rate in the Current Population Survey. In: Koch, A. and Porst, R. (eds.). Nonresponse in Survey Research. Proceedings of the Eighth International Workshop on Household Survey Nonresponse 24-26 September 1997. Mannheim: ZUMA, pp. 45-54

Zabel, J. E. (1994). A Comparison of Attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation. Unpublished manuscript. U. S. Census Bureau, Washington, D.C.